A Family of Joint Probability Models for Cost and Schedule Uncertainties

Paul R. Garvey
The MITRE Corporation, Bedford, Massachusetts

27th Annual Department of Defense Cost Analysis Symposium
September 1993

Abstract - This paper discusses a family of analytical models from which the joint probability of total program cost and schedule can be calculated, analyzed, and presented to decision-makers. Specifically, the classical bivariate normal and two lesser known joint distributions, the normal-lognormal and the bivariate lognormal distributions are discussed. Experiences from Monte Carlo simulations suggest that this family of bivariate distributions are candidate models for computing joint and conditional cost and schedule probabilities. In particular, the discussion on the normal-lognormal distribution as a joint cost-schedule probability model extends research on the applicability of the bivariate lognormal presented last year*.

Joint probability distributions enable analysts and decision-makers to determine joint and conditional probabilities of the types

$$P(\text{Cost} \le a \text{ and Schedule} \le b)$$

$$P(\text{Cost} \le a \mid \text{Schedule} = b)$$

at the program level. Probability statements such as these have not been a common product of cost uncertainty analyses, and the development of analytical methods that produce these probabilities is an advance in the state of the practice.

General formulas for the correlation functions, conditional distributions, and moments of the family of joint probability distributions described in this paper are provided. Cost analysis applications are presented to illustrate the use of these distributions in a practical context.

* Garvey, P. R., Taub, A. E., *A Joint Probability Model for Cost and Schedule Uncertainties*, The MITRE Corporation, presented at the 26th ADODCAS, September 1992.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1993	ADODCAS 1993

4. TITLE AND SUBTITLE
A FAMILY OF JOINT PROBABILITY MODELS FOR COST AND SCHEDULE UNCERTAINTIES

5. FUNDING NUMBERS

6. AUTHOR(S)
Paul R. Garvey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The MITRE Corporation
202 Burlington Road
Bedford MA
(617) 271-8234

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Annual Department of Defense Cost Analysis Symposium Paper

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Statement A: Approved for Public Release; Distribution is Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
See Over

DTIC QUALITY INSPECTED 1

14. SUBJECT TERMS
Probability Models
Cost and Schedule
Risk

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	UL

A FAMILY OF JOINT PROBABILITY MODELS FOR
COST AND SCHEDULE UNCERTAINTIES

This paper proposes a family of analytical models from which the joint probability of total program cost and schedule can be described, analyzed, and presented to decision-makers. Specifically, the classical bivariate normal and two lesser known joint distributions, the normal-lognormal and the bivariate lognormal distributions are discussed. Experiences from Monte Carlo simulations suggest that this family of bivariate distributions are candidate models for computing joint and conditional cost and schedule probabilities. In particular, the discussion on the normal-lognormal distribution as a joint cost-schedule probability model extends research on the applicability of the bivariate lognormal presented last year[1].

Joint probability distributions enables analysts and decision-makers to conduct tradeoffs on joint and conditional probabilities of the types

$$P(\text{Cost} \le a \text{ and Schedule} \le b)$$
$$P(\text{Cost} \le a \mid \text{Schedule} = b)$$

at the program level. Probability statements such as these have not been a common product of cost uncertainty analyses, and the development of analytical methods that produce these probabilities is an advance in the state of the practice.

General formulas for the correlation functions, conditional distributions, and moments of the family of joint probability distributions *described in* this paper are provided.

Paul R. Garvey
The MITRE Corporation
202 Burlington Road
Bedford MA
(617) 271-8234

[1] Garvey, P. R., Taub A.E., *A Joint Probability Model for Cost and Schedule Uncertainties*. The MITRE Corporation, presented at the 26th ADODCAS, September 1992.

A Family of Joint Probability Models for Cost and Schedule Uncertainties

Paul R. Garvey
The MITRE Corporation, Bedford, Massachusetts

27th Annual Department of Defense Cost Analysis Symposium
September 1993

Abstract - This paper discusses a family of analytical models from which the joint probability of total program cost and schedule can be calculated, analyzed, and presented to decision-makers. Specifically, the classical bivariate normal and two lesser known joint distributions, the normal-lognormal and the bivariate lognormal distributions are discussed. Experiences from Monte Carlo simulations suggest that this family of bivariate distributions are candidate models for computing joint and conditional cost and schedule probabilities. In particular, the discussion on the normal-lognormal distribution as a joint cost-schedule probability model extends research on the applicability of the bivariate lognormal presented last year*.

Joint probability distributions enable analysts and decision-makers to determine joint and conditional probabilities of the types

$$P(\text{Cost} \le a \text{ and Schedule} \le b)$$
$$P(\text{Cost} \le a \,|\, \text{Schedule} = b)$$

at the program level. Probability statements such as these have not been a common product of cost uncertainty analyses, and the development of analytical methods that produce these probabilities is an advance in the state of the practice.

General formulas for the correlation functions, conditional distributions, and moments of the family of joint probability distributions described in this paper are provided. Cost analysis applications are presented to illustrate the use of these distributions in a practical context.

* Garvey, P. R., Taub, A. E., *A Joint Probability Model for Cost and Schedule Uncertainties*, The MITRE Corporation, presented at the 26th ADODCAS, September 1992.

I. INTRODUCTION

When cost and schedule uncertainty analyses are presented to decision-makers, questions asked with increased urgency are: What is the likelihood of achieving both the estimated cost and schedule? What is the chance of exceeding the most likely cost for a given schedule? How are cost reserve recommendations driven by schedule uncertainties? Questions such as these are not readily answered by current analysis tools and techniques. This paper advances the state of the practice by presenting a family of probability models designed to specifically address these and related questions.

During the past thirty years, uncertainty analysis methods have applied univariate probability theory to generate separate distributions of a program's estimated cost and schedule. These distributions provide decision-makers with probabilities such as

$$P(\text{Cost} \le a) \tag{1}$$

and

$$P(\text{Schedule} \le b) \tag{2}$$

Although it has long been recognized that cost and schedule are interdependent, little has been adopted from multivariate theory to combine cost and schedule probability distributions. A multivariate model that combines these distributions would provide decision-makers joint and conditional probabilities such as

$$P(\text{Cost} \le a \text{ and Schedule} \le b) \tag{3}$$

and

$$P(\text{Cost} \le a \mid \text{Schedule} = b) \tag{4}$$

In practice, probabilities in (3) and (4) are of greater interest to decision-makers than those in (1) and (2). This is particularly true in the early planning phases of a program where critical cost and schedule decisions are made.

This paper presents the cost analysis community with a family of multivariate probability models from which joint and conditional distributions of program cost and schedule can be developed. Specifically, the classical bivariate normal and two lesser known joint distributions, the bivariate normal-lognormal and the bivariate lognormal distributions are discussed. The development of the bivariate normal-lognormal distribution, and its application in the cost analysis domain, extends research [1] on the utility of joint probability models for program cost and schedule. Appendixes are provided that document the properties of the bivariate normal-lognormal and bivariate lognormal models. Properties of the bivariate normal are well known and can be found in [2].

The models proposed in this paper are analytical. They facilitate computing the probabilities defined in (3) and (4). Furthermore, the means and variances from univariate cost and schedule distributions developed from traditional techniques can be used to specify the parameters of a bivariate probability model assumed for a given program. Thus, no new or additional analysis effort is required to define and apply these models. Lastly, these models explicitly incorporate correlation between cost and schedule on a given program - an issue not well addressed in current analysis methodology.

Formulas are developed that combine program cost and schedule probability distributions to produce conditional cost distributions for a given set of schedules. For example, in figure 1 the curves with solid lines are unconditional distributions of a program's cost and schedule. They represent probabilities of the types in (1) and (2). Formulas are provided in section III that incorporate statistics from these distributions to form conditional cost distributions for a given set of schedules. The curves with dashed lines in figure 1 are examples of conditional cost distributions. Conditional distributions represent probabilities of the type defined

in (4). They are an effective way of directly linking a given schedule to cost recommendations tied to a specified level of confidence.

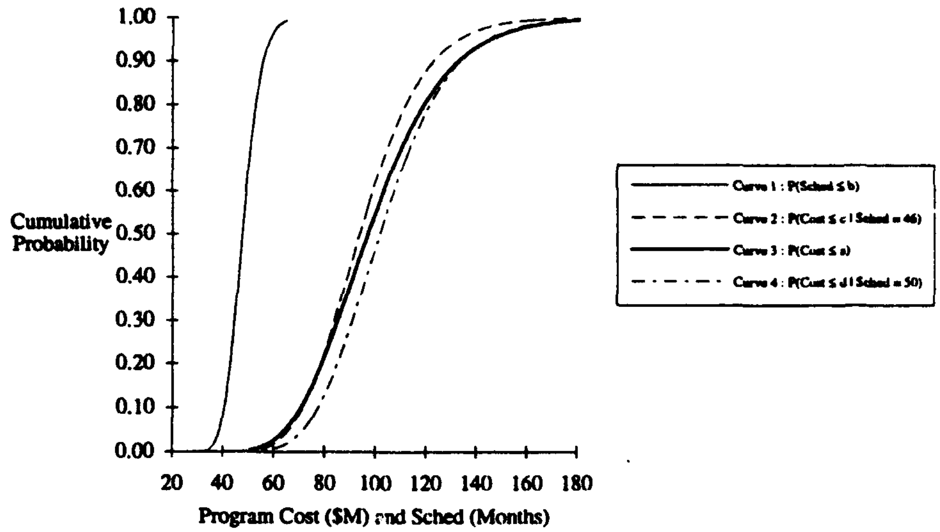

Figure 1. Illustrative Probability Distributions of Program Cost and Schedule

In summary, this paper advances current practice by offering a family of multivariate models that produce joint and conditional probabilities of program cost and schedule. Cost analysis applications that illustrate the mathematical theory are presented.

II. JOINT PROBABILITY MODELS AND THEIR SUPPORT TO RISK MANAGEMENT

This section briefly defines the domain of the models described in this paper and how they provide analyses that supports risk management decisions. An overview of the risk management decision space is discussed. A context for how these models are defined within that space is provided. Finally, the types of uncertainties that can be quantified by these models are discussed.

A. Models and the Risk Management Decision Space

Cost, schedule, and performance objectives define the domain of the program manager's decision space. A characterization of this space is presented in figure 2. Illustrated is how user expectations on cost, schedule, and performance are often at odds with what can actually be delivered. Risk is introduced when expectations in any of these dimensions push what is technically and/or economically feasible. Managing risk, then, is managing the inherent contention that exists within each axis and across all three. The goal of risk management is to identify cost, schedule, and performance risks early such that control on either axis is not lost, or that the consequences of risk mitigating actions on all three axes are well understood.

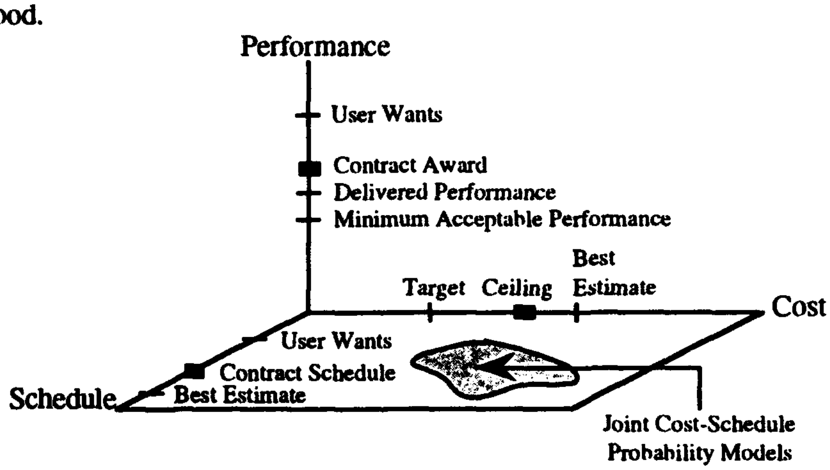

Figure 2. Models and the Risk Management Decision Space

The joint probability models presented in section III operate in two of the three dimensions shown in figure 2. They are bivariate in the cost and schedule plane. When applying these models it is assumed that a system architecture has been defined, or hypothesized, by the engineering team. It is also assumed that this

architecture will achieve the performance requirements as they are specified or understood at that time.

Thus, for a given system architecture, the uncertainties captured by these models are those associated with its technical definition and resource (cost-schedule) estimation approaches. Examples of technical definition uncertainties include estimating the amount of new code to develop, or the number of custom microchips to design and fabricate, or the delivered weight of a newly designed antenna pedestal. Examples of resource estimation uncertainties include errors in cost models and cost methodologies, uncertainties in estimated activity durations, economic uncertainties as they affect the cost of technology (e.g., inflation/deflation), or uncertainties in the data (e.g., labor rates, productivity rates) used to develop the resource estimates.

Once a joint cost-schedule probability model has been developed on a program, analyses can be conducted to support risk management decisions. These include

Baselining Program Cost and Schedule Risk - For a given system architecture, acquisition strategy, and resource estimation approach baseline probability distributions of a program's cost and schedule can be developed. These distributions can be periodically updated as the program's uncertainties change across the acquisition milestones. Generating these distributions supports estimating a program cost and schedule that simultaneously have a specified probability of not being exceeded. This distribution also provides program managers with an assessment of the likelihood of achieving a budgeted or proposed cost and schedule or cost for a given schedule.

Estimating Reserves - An analytical basis for estimating cost reserve for a given schedule, or set of schedules, as a function of the uncertainties specific to a given system can be developed. Sensitivity analyses can be conducted to assess how reserve levels are affected by changes in specific program risks. In addition, the relationship between the amount of reserve to recommend for a given level of confidence and a given schedule can be examined.

Conducting Risk Mitigation Tradeoff Analyses - Models can be developed to study the payoff of implementing specific risk amelioration activities (e.g., prototyping) on reducing the estimated cost and schedule variances. For instance, suppose that the cost and schedule variances of a program are driven by uncertainty in the amount of new code to develop. Using these models, it can be determined whether investing in a prototyping effort markedly reduces this variance, and, therefore, lessen the cost and schedule reserves estimated for the program.

III. A FAMILY OF JOINT PROBABILITY MODELS FOR COST AND SCHEDULE

This section presents a family of joint probability models for program cost and schedule. Specifically, the bivariate normal, the bivariate normal-lognormal, and the bivariate lognormal models are described. These models are candidate theoretical distributions that might be assumed by an analyst when joint or conditional cost-schedule probabilities are needed.

The models described herein have two features desirable for cost analysis applications. First, they directly incorporate correlation between cost and schedule on a given program. Second, they have the property that their marginal distributions* are either both normal, normal and lognormal, or both lognormal. This is reflective of the distributions frequently observed in Monte Carlo simulations [1,3] of program cost and schedule.

The following briefly describes these models. A cost analysis application is also presented. Throughout the remainder of this paper the random variables X_1 and X_2 will denote program cost and schedule, respectively.

---

* If X_1 and X_2 are random variables with joint probability distribution $F_{X_1,X_2}(x_1,x_2)$, then the marginal distributions of X_1 and X_2 are the individual probability distributions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$, respectively.

A. The Classical Bivariate Normal Model

This section presents the classical bivariate normal and summarizes its major characteristics. An important feature of this model are its marginal distributions, which are both normal.

In cost analyses, normal distributions can arise when program cost is the sum of many uncorrelated WBS cost element random variables. Normal distributions can also occur in schedule analyses. For instance, program schedule is approximately normal if it is the sum of many independent activities in a network. Thus, if normal distributions characterize a program's cost and schedule distributions, then the bivariate normal could serve as an assumed model of their joint distribution.

A.1 Model Definition

Suppose we have two random variables $Y_1 = X_1$ and $Y_2 = X_2$ where X_1 and X_2 are defined on $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$. If Y_1 and Y_2 each have a normal distribution then, for $i = 1, 2$, the mean and variance of Y_i are

$$E(Y_1) = \mu_{Y_1} = \mu_{X_1} = \mu_1 \tag{5}$$

$$E(Y_2) = \mu_{Y_2} = \mu_{X_2} = \mu_2 \tag{6}$$

$$Var(Y_1) = \sigma^2_{Y_1} = \sigma^2_{X_1} = \sigma^2_1 \tag{7}$$

$$Var(Y_2) = \sigma^2_{Y_2} = \sigma^2_{X_2} = \sigma^2_2 \tag{8}$$

Assume that the pair

$$(X_1, X_2) \sim \text{Bivariate } N((\mu_1, \mu_2), (\sigma^2_1, \sigma^2_2, \rho_{1,2})) \tag{9}$$

is a bivariate normal distribution with density function [2]

$$f_{X_1,X_2}(x_1,x_2)=\begin{cases}\dfrac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho_{1,2}^2}}e^{-\frac{1}{2}w} & -\infty<x_1<\infty\,,\ -\infty<x_2<\infty \\ 0 \quad \text{otherwise} \end{cases} \tag{10}$$

where

$$w=\frac{1}{1-\rho_{1,2}^2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho_{1,2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

and μ_i and σ_i^2 (i = 1, 2) are given by equations (5) through (8). The correlation term $\rho_{1,2}$ in equation (10) is

$$\rho_{1,2}=\rho_{Y_1,Y_2}=\rho_{X_1,X_2} \tag{11}$$

If two continuous random variables X_1 and X_2 have a bivariate normal distribution, then

$$P(a_1\le X_1\le b_1 \text{ and } a_2\le X_2\le b_2)$$

is given by

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1} f_{X_1,X_2}(x_1,x_2)\,dx_1dx_2 \tag{12}$$

where, in this case, f is defined by equation 10.

A.II Model Characteristics

A characteristic of the bivariate normal distribution is that the distribution of X_1 and the distribution of X_2 are each univariate normal. These are the marginal distributions and they are given by

$$f_1(x_1)=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{1}{2}\left[(x_1-\mu_1)^2/\sigma_1^2\right]} \tag{13}$$

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left[(x_2-\mu_2)^2/\sigma_2^2\right]} \tag{14}$$

Important tradeoffs in cost analysis often involve assessing the impact that a given set of schedules have on the likelihood that program cost will not exceed a prescribed threshold. To make these assessments, the conditional probability distribution is needed. Conditional distributions provide probabilities of the type

$$P(X_1 \le a \mid X_2 = b)$$

If two continuous random variables X_1 and X_2 have a bivariate normal distribution, then the conditional probability density function of X_1 given $X_2 = x_2$, denoted by $f_{X_1|x_2}(x_1)$, is normally distributed [2]. That is

$$X_1 \mid x_2 \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(x_2-\mu_2), \sigma_1^2(1-\rho_{1,2}^2)) \tag{15}$$

similarly

$$X_2 \mid x_1 \sim N(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1), \sigma_2^2(1-\rho_{1,2}^2)) \tag{16}$$

From (15) and (16) the conditional means and variances are

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(x_2-\mu_2) \tag{17}$$

$$E(X_2|x_1) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1) \tag{18}$$

and

$$\mathrm{Var}(X_1|x_2) = \sigma_1^2(1-\rho_{1,2}^2) \tag{19}$$

$$\mathrm{Var}(X_2|x_1) = \sigma_2^2(1-\rho_{1,2}^2) \tag{20}$$

Section D will illustrate how conditional distributions are developed to provide decision-makers with cost-schedule probabilities such as

$$P(\text{Cost} \le a \mid \text{Schedule} = b)$$

B. A Bivariate Normal-LogNormal Model

This section presents a bivariate normal-lognormal model and summarizes its major characteristics. Unlike the classical bivariate normal, very little exists in the statistical literature on this model. Furthermore, the few references [4,5] that discuss the bivariate normal-lognormal define it by parameters not computed in cost-schedule uncertainty analyses. Thus, an alternative form of the model than that offered in [4,5] was created specifically for this application context. Its form is a direct extension of the bivariate normal presented in section A. A theoretical discussion of this model is provided in appendix A.

An important feature of the bivariate normal-lognormal model is its marginal distributions. One marginal is normal and the other is lognormal (refer to appendix A). Section A described circumstances that give rise to normal marginal distributions for cost or schedule. Under certain conditions, lognormal distributions can characterize simulated program cost and schedule distributions [1,3]. For instance, it is frequently observed that the lognormal approximates the simulated distribution of program cost if it is the sum of many positively correlated cost element random variables. Similarly, program schedule can also tend towards lognormality if it is the sum of many positively correlated schedule activities.

These observations stem from practical experiences with Monte Carlo simulations. They are not theoretical findings. Exceptions to these observations can be created. However, if a situation arises where normal and lognormal distributions characterize a program's cost and schedule distributions, then the bivariate normal-lognormal could serve as an assumed model of their joint distribution.

B.1 Model Definition

Suppose we have two random variables $Y_1 = X_1$ and $Y_2 = \ln X_2$ where X_1 and X_2 are defined on $-\infty < x_1 < \infty$ and $0 < x_2 < \infty$. If Y_1 and Y_2 each have a normal distribution then, for i =1, 2, the mean and variance of Y_i [1] are

$$E(Y_1) = \mu_{Y_1} = \mu_{X_1} = \mu_1 \tag{21}$$

$$Var(Y_1) = \sigma^2_{Y_1} = \sigma^2_{X_1} = \sigma^2_1 \tag{22}$$

$$E(Y_2) = \mu_{Y_2} = \mu_2 = \frac{1}{2}\ln\left[\frac{(\mu_{X_2})^4}{(\mu_{X_2})^2 + \sigma^2_{X_2}}\right] \tag{23}$$

$$Var(Y_2) = \sigma^2_{Y_2} = \sigma^2_2 = \ln\left[\frac{(\mu_{X_2})^2 + \sigma^2_{X_2}}{(\mu_{X_2})^2}\right] \tag{24}$$

Assume that the pair

$$(X_1, X_2) \sim \text{Bivariate NLogN}((\mu_1, \mu_2), (\sigma^2_1, \sigma^2_2, \rho_{1,2})) \tag{25}$$

is a bivariate normal-lognormal distribution with density function

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \dfrac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2_{1,2}}\; x_2} e^{-\frac{1}{2}w} & -\infty < x_1 < \infty,\ 0 < x_2 < \infty \\ 0 \quad \text{otherwise} \end{cases} \tag{26}$$

where

$$w = \frac{1}{1-\rho^2_{1,2}}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{1,2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{\ln x_2-\mu_2}{\sigma_2}\right) + \left(\frac{\ln x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

and μ_i and σ^2_i (i = 1,2) are given by equations (21) through (24). From appendix A (theorem A-1) the correlation term $\rho_{1,2}$ in equation (26) is

$$\rho_{1,2} = \rho_{Y_1,Y_2} = \rho_{X_1,\ln X_2} = \rho_{X_1,X_2} \frac{(e^{\sigma_2^2} - 1)^{1/2}}{\sigma_2} \tag{27}$$

Equation 27 can be found in P. T. Yuan (1933) [6]. Proofs have been developed by P. H. Young* and the author (refer to appendix A).

If two continuous random variables X_1 and X_2 have a bivariate normal-lognormal distribution, then

$$P(a_1 \le X_1 \le b_1 \text{ and } a_2 \le X_2 \le b_2)$$

is given by

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 \tag{28}$$

where, in this case, f is defined by equation 26

B.II Model Characteristics

A characteristic of the bivariate normal-lognormal distribution is that the distribution of X_1 is normal and the distribution of X_2 is lognormal. These marginal distributions are given by

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left[(x_1 - \mu_1)^2 / \sigma_1^2\right]} \tag{29}$$

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2 / \sigma_2^2\right]} \tag{30}$$

---

* P. H. Young [The Aerospace Corporation, 1992] developed a proof of (27). His approach used the definition of $E(X_1 X_2)$ to determine $\rho_{1,2}$. The proof of (27) presented in appendix A is a slight variation of Young's approach. The formula for $\rho_{1,2}$ is established directly from the definition of the $Cov(X_1 X_2)$ instead of $E(X_1 X_2)$. Recall that $Cov(X_1 X_2) = E(X_1 X_2) - E(X_1)E(X_2)$.

The conditional probability density function of X_1 given $X_2 = x_2$, denoted by $f_{X_1|x_2}(x_1)$, is normally distributed (refer to appendix A, theorem A-3). That is

$$X_1|x_2 \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2)) \tag{31}$$

similarly

$$X_2|x_1 \sim \mathrm{LogN}(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1 - \mu_1), \sigma_2^2(1-\rho_{1,2}^2)) \tag{32}$$

From appendix A (theorem A-4) the conditional means and variances are

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2) \tag{33}$$

$$E(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1) + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)} \tag{34}$$

and

$$Var(X_1|x_2) = \sigma_1^2(1-\rho_{1,2}^2) \tag{35}$$

$$Var(X_2|x_1) = e^{2(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1))} e^z(e^z - 1) \tag{36}$$

where $z = \sigma_2^2(1-\rho_{1,2}^2)$

Section D will illustrate how conditional distributions from the normal-lognormal model are developed to provide decision-makers with cost-schedule probabilities such as

$$P(\text{Cost} \le a \,|\, \text{Schedule} = b)$$

C. The Bivariate LogNormal Model

This section presents the bivariate lognormal model and summarizes its major characteristics. Research presented last year [1] demonstrated that, under certain conditions, this model can approximate a program's cost-schedule joint and

conditional probabilities. The bivariate lognormal has a number of characteristics desirable for cost analysis applications. Among these are

- It is bounded by zero (cost and schedule cannot become negative),
- Its marginal and conditional distributions are lognormal, which is frequently observed in Monte Carlo simulations of program cost and schedule [1,3].

C.1 Model Definition

Suppose we have two random variables $Y_1 = \ln X_1$ and $Y_2 = \ln X_2$ where X_1 and X_2 are defined on $0 < x_1 < \infty$ and $0 < x_2 < \infty$. If Y_1 and Y_2 each have a normal distribution then, for i = 1, 2, the mean and variance of Y_i [1] are

$$E(Y_i) = \mu_{Y_i} = \mu_i = \frac{1}{2}\ln\left[\frac{(\mu_{X_i})^4}{(\mu_{X_i})^2 + \sigma_{X_i}^2}\right] \tag{37}$$

$$\mathrm{Var}(Y_i) = \sigma_{Y_i}^2 = \sigma_i^2 = \ln\left[\frac{(\mu_{X_i})^2 + \sigma_{X_i}^2}{(\mu_{X_i})^2}\right] \tag{38}$$

Assume that the pair

$$(X_1, X_2) \sim \text{Bivariate LogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2})) \tag{39}$$

is a bivariate lognormal distribution with density function

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \dfrac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho_{1,2}^2}\,x_1 x_2} e^{-\frac{1}{2}w} & 0 < x_1 < \infty, 0 < x_2 < \infty \\ 0 \quad \text{otherwise} & \end{cases} \tag{40}$$

where

$$w = \frac{1}{1-\rho_{1,2}^2}\left\{\left(\frac{\ln x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho_{1,2}\left(\frac{\ln x_1 - \mu_1}{\sigma_1}\right)\left(\frac{\ln x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\ln x_2 - \mu_2}{\sigma_2}\right)^2\right\}$$

and μ_i and σ_i^2 ($i = 1, 2$) are given by equations (37) and (38). From appendix B (theorem B-2), the correlation term $\rho_{1,2}$ in equation (40) is

$$\rho_{1,2} = \frac{1}{\sigma_1\sigma_2}\ln\left[1+\rho_{X_1,X_2}\sqrt{e^{\sigma_1^2}-1}\sqrt{e^{\sigma_2^2}-1}\right] \tag{41}$$

If two continuous random variables X_1 and X_2 have a bivariate lognormal distribution, then

$$P(a_1 \le X_1 \le b_1 \text{ and } a_2 \le X_2 \le b_2)$$

is given by

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1} f_{X_1,X_2}(x_1,x_2)\,dx_1dx_2 \tag{42}$$

where, in this case, f is defined by equation 40.

C.II Model Characteristics

A characteristic of the bivariate lognormal distribution is that the distribution of X_1 and the distribution of X_2 are each lognormal. These marginal distributions are given by

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1 x_1}e^{-\frac{1}{2}\left[(\ln x_1-\mu_1)^2/\sigma_1^2\right]} \tag{43}$$

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2 x_2}e^{-\frac{1}{2}\left[(\ln x_2-\mu_2)^2/\sigma_2^2\right]} \tag{44}$$

The conditional probability density function of X_1 given $X_2 = x_2$, denoted by $f_{X_1|x_2}(x_1)$, is lognormally distributed (refer to appendix B, theorem B-4). That is

$$X_1|\,x_2 \sim \text{LogN}(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2)) \tag{45}$$

similarly $$X_2 | x_1 \sim \text{LogN}(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1), \sigma_2^2(1-\rho_{1,2}^2)) \quad (46)$$

From appendix B (theorem B-5) the conditional means and variances are

$$E(X_1 | x_2) = x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2 + \frac{1}{2}\sigma_1^2(1-\rho_{1,2}^2)} \quad (47)$$

$$E(X_2 | x_1) = x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1 + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)} \quad (48)$$

and

$$\text{Var}(X_1 | x_2) = x_2^{2\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{2(\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2)} e^{z^*}(e^{z^*} - 1) \quad (49)$$

$$\text{Var}(X_2 | x_1) = x_1^{2\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{2(\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1)} e^{z}(e^{z} - 1) \quad (50)$$

where

$$z = \sigma_2^2(1-\rho_{1,2}^2) \quad \text{and} \quad z^* = \sigma_1^2(1-\rho_{1,2}^2)$$

Section D will illustrate how conditional distributions from the bivariate lognormal model are developed to provide decision-makers with cost-schedule probabilities such as

$$P(\text{Cost} \leq a \,|\, \text{Schedule} = b)$$

D. A Cost Analysis Application

The following illustrates, from a cost analysis perspective, how the joint probability models described in this paper can be defined and applied. Recall that the random variables X_1 and X_2 denote program cost and schedule, respectively.

Suppose that the cumulative probability distributions of a program's cost and schedule are given in figure 3. Furthermore, suppose that these distributions have means and variances given in table 1 and that the correlation between program cost and schedule, ρ_{X_1,X_2}, was 0.50.

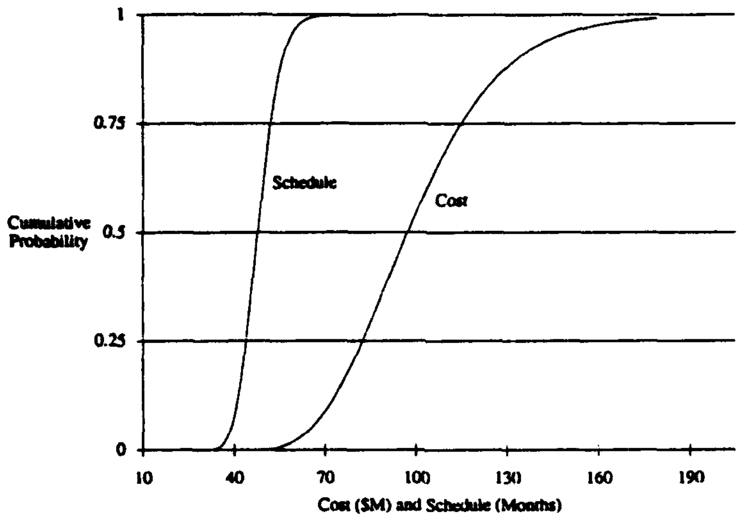

Figure 3. Probability Distributions for Program Cost and Schedule

Table 1. Program Cost and Schedule Statistics

i	X_i	μ_{X_i}	$\sigma^2_{X_i}$	σ_{X_i}
1	Cost ($M)	100	625	25
2	Schedule (Months)	48	36	6
	Correlation ρ_{X_1,X_2}		0.50	

Using these data, the parameters of the three joint probability models described in this paper can be determined. These parameters are summarized in table 2.

From table 2 and (9), (25), and (39) we can write

$$(X_1, X_2) \sim \text{Bivariate N}((100, 48), (625, 36, 0.50)) \tag{51}$$

$$(X_1, X_2) \sim \text{Bivariate NLogN}((100, 3.863), (625, 0.0155, 0.502)) \tag{52}$$

$$(X_1, X_2) \sim \text{Bivariate LogN}((4.575, 3.863), (0.0606, 0.0155, 0.506)) \tag{53}$$

if the pair (X_1, X_2) is an assumed bivariate normal, an assumed bivariate normal-lognormal, or an assumed bivariate lognormal.

Table 2. Joint Probability Model Parameter Set

i	μ_i	σ_i^2	σ_i
Bivariate Normal Model			
1	100 *(Eqt 5)*	625 *(Eqt 7)*	25
2	48 *(Eqt 6)*	36 *(Eqt 8)*	6
	Correlation $\rho_{1,2}$	0.50 *(Eqt 11)*	
Bivariate Normal-LogNormal Model			
1	100 *(Eqt 21)*	625 *(Eqt 22)*	25
2	3.863 *(Eqt 23)*	0.0155 *(Eqt 24)*	0.124
	Correlation $\rho_{1,2}$	0.502 *(Eqt 27)*	
Bivariate LogNormal Model			
1	4.575 *(Eqt 37)*	0.0606 *(Eqt 38)*	0.246
2	3.863 *(Eqt 37)*	0.0155 *(Eqt 38)*	0.124
	Correlation $\rho_{1,2}$	0.506 *(Eqt 41)*	

Once the joint distribution of (X_1, X_2) has been specified, joint cost-schedule probabilities can be computed. For instance, if (X_1, X_2) is assumed to be bivariate normal-lognormal with parameters given in (52), then the probability

$$P(\$100M \leq X_1 \leq \$130M \text{ and } 45 \text{ months} \leq X_2 \leq 55 \text{ months})$$

is

$$\int_{45}^{55}\int_{100}^{130} f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2 = 0.245$$

In this case, the integrand $f_{X_1,X_2}(x_1,x_2)$ is defined by equation 26. Joint cost-schedule probabilities for other regions of interest are determined in a similar manner.

D.1 Conditional Cost-Schedule Distributions

Cost analyses must often focus on assessing the impact that a given set of schedules have on the likelihood that an estimated program cost will not be exceeded. To make these assessments the conditional distribution of program cost is needed. Stated previously, conditional distributions produce probabilities of the type

$$P(\text{Cost} \le a | \text{Schedule} = b)$$

Computing a conditional probability requires the joint cost-schedule probability distribution. Recall the following joint cost-schedule distributions formed from the data in table 1.

$$(X_1,X_2) \sim \text{Bivariate N}((100,48),(625,36,0.50))$$
$$(X_1,X_2) \sim \text{Bivariate NLogN}((100,3.863),(625,0.0155,0.502))$$
$$(X_1,X_2) \sim \text{Bivariate LogN}((4.575,3.863),(0.0606,0.0155,0.506))$$

From equations 15, 31, and 45, and table 2, the conditional distributions of cost $X_1|x_2$ for a given schedule x_2 can be formed. In this case, the conditional distribution of $X_1|x_2$ if the pair (X_1,X_2) is an assumed bivariate normal, an assumed bivariate normal-lognormal, or an assumed bivariate lognormal are, respectively

$$X_1|x_2 \sim N(100+2.083(x_2-48), 468.75) \tag{54}$$

$$X_1|x_2 \sim N(100+101.21(\ln x_2-3.863), 467.5) \tag{55}$$

$$X_1|x_2 \sim \text{LogN}(4.575+1.0038(\ln x_2-3.863), 0.045) \tag{56}$$

Using the lognormal distribution given by (56), conditional cumulative cost distributions given x_2 equal to 50, 55, and 60 months are shown in figure 4. Table 3 summarizes the statistics, computed from theorems B-4 and B-5 (refer to appendix B), of these distributions.

From figure 4, the impact of a given schedule on the probability distribution of cost can be determined. Observe that as x_2 increases, the cumulative conditional cost distributions become "lazier." Shown in table 3, the increased laziness reflects a growth of nearly 2 million dollars in $\sigma(X_1|x_2)$ for each 5 month increase in x_2.

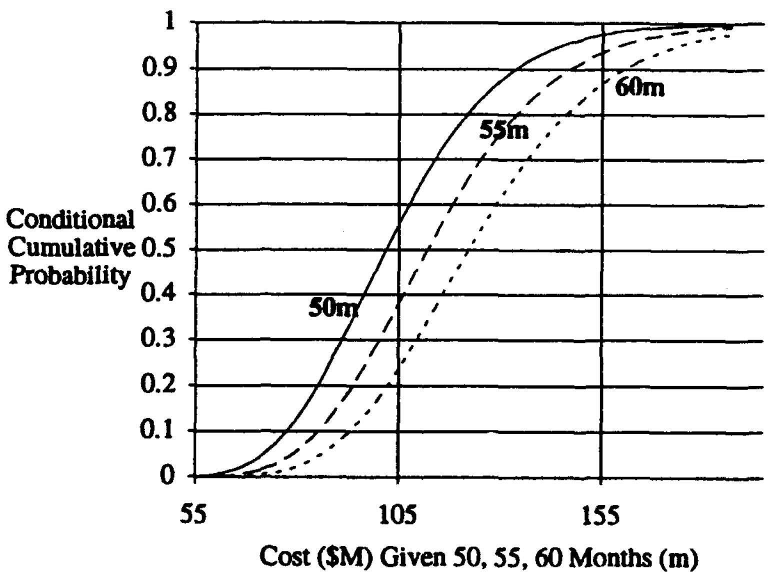

Figure 4. A Family of Conditional LogNormal Cost Distributions
(Equation 56 with $x_2 = 50, 55, 60$ months)

By definition, the conditional median cost occurs with probability

$$P(\text{Cost} \leq a | \text{Schedule} = x_2) = 0.50$$

where, in table 3, a is equal to 102, 112, and 122 million dollars given x_2 equal to 50, 55, and 60 months, respectively. This reflects the cost that is equally likely to overrun or underrun for a given schedule x_2. For this specific example, $\text{Median}(X_1|x_2)$ increases by 10 million dollars for every 5 month increase in x_2. A similar increase is seen in the conditional mean cost $E(X_1|x_2)$.

Table 3. Statistics from the Conditional Distributions in Figure 4

Given x_2 (Months)	$\text{Median}(X_1\|x_2)$ ($M)	$E(X_1\|x_2)$ ($M)	$\sigma(X_1\|x_2)$ ($M)
50	102	104	22.4
55	112	115	24.6
60	122	125	26.9

Linkages such as these between cost uncertainty and schedule can be made through the use of conditional distributions. In the early stages of a program, conditional distributions provide decision-makers valuable insight into the likelihood of achieving cost and schedule goals.

IV. SUMMARY COMMENTS

The family of joint probability models described in this paper provides an analytical basis for computing joint and conditional cost-schedule probabilities. Selection of a particular model is guided by the marginal distributions it produces. For example, if the individual distributions of X_1 and X_2 are observed to be normal and lognormal, then the bivariate normal-lognormal model (25) might be chosen for the joint distribution of the pair (X_1, X_2). This is because the bivariate normal-lognormal model produces normal and lognormal marginal distributions. However, it must be viewed that choosing this particular model makes an

assumption about how the pair of random variables (X_1, X_2) is jointly distributed. The true joint distribution of (X_1, X_2) cannot be uniquely determined from only the individual distributions of X_1 and X_2.

A parameter required by the models in this paper is the correlation between program cost and schedule. Although this is a difficult parameter to estimate, approaches include deriving it from an historical cost-schedule database (one such correlation coefficient can be found in reference 1) or computing it using values sampled from a simulation of the cost-schedule estimating relationships established for a particular program. Subjective assessments might be used in the absence of a database or a cost-schedule simulation model.

An important consideration regarding the models herein is that they do not reflect the causal impact that schedule compression or extension has on cost. These models treat cost and schedule as correlated random variables whose range of values are reflected by their marginal distributions. These ranges result from quantifying the uncertainties associated with a specific technical baseline and the cost-schedule estimation approaches. Unrealistically compressing or extending schedule (due to missed milestones or program re-plans) can incur increased cost. In these circumstances a reassessment of the system's cost-schedule risk is warranted.

The utility of joint probability models is enabling analysts to present decision-makers with an integrated view of cost and schedule uncertainties. In developing the model parameters for a specific system, elements that significantly contribute to cost and schedule risk are identified. This fosters early recognition of inadequately specified requirements and permits risk mitigating decisions to be made early in the system definition phases. For the program manager, joint probability models reveal the simultaneous cost and schedule impacts of uncertainties in such areas as

technical realism, the soundness of the acquisition approach, the reasonableness of the schedule, and the accuracy of the cost and schedule estimation methodologies.

In an environment of limited funds and increasingly challenging schedules, it is incumbent upon analysts to continually examine affordability concerns relative to the likelihood of jointly meeting cost and schedule $P(\text{Cost} \leq a \text{ and Schedule} \leq b)$ and cost for a given schedule $P(\text{Cost} \leq a \mid \text{Schedule} = b)$ against specific tradeoffs in system requirements, acquisition strategies, and post-development support. Enabling options to be explored that offer decision-makers economically sound and risk mitigating choices throughout the life of a program is the cost analyst's aim and opportunity. Models and methodologies are tools that provide a means to that end.

ACKNOWLEDGMENTS

The author gratefully acknowledges the technical review of this paper by Mr. David N. Lam, Dr. Neal D. Hulkower, Mr. Charles M. Plummer, and Dr. Karen W. Pullen of The MITRE Corporation.

REFERENCES

[1] Garvey, P. R., Taub, A. E., *A Joint Probability Model for Cost and Schedule Uncertainties*, The MITRE Corporation, presented at the 26th Annual Department of Defense Cost Analysis Symposium, September 1992.

[2] Olkin, I., Gleser, L. J., Derman, C., Probability Models and Applications, Macmillan Publishing Co., Inc., New York, 1980.

[3] Abramson, R. L., Young, P. H., *FRISKEM-Formal Risk Evaluation Methodology*, presented at the ORSA/TIMS Joint National Meeting, Philadelphia, PA, October 29-31, 1990, The Aerospace Corporation, Los Angeles, CA.

[4] Patil, G. P., Boswell, M. T., Ratnaparkhi, M. V., Roux, J. J. J., Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, International Co-operative Publishing House, Fairland, Maryland, 1984.

[5] Crofts, A. E., *An Investigation of the Bivariate Normal LogNormal Distributions*, Technical Report Number 32, Themis Contract, Southern Methodist University, 21 April 1969.

[6] Yuan, P. T., *On the Logarithmic Frequency Distribution and the Semi-logarithmic Correlation Surface*, Annals of Mathematical Statistics, 4, 30-74, 1933.

Appendix A

Theoretical Aspects of the Normal-LogNormal Joint Probability Distribution

Let $Y_1 = X_1$ and $Y_2 = \ln X_2$ where X_1 and X_2 are random variables defined on $-\infty < x_1 < \infty$ and $0 < x_2 < \infty$. If Y_1 and Y_2 each have a normal distribution then

$$E(Y_1) = \mu_{Y_1} = \mu_{X_1} = \mu_1$$

$$Var(Y_1) = \sigma_{Y_1}^2 = \sigma_{X_1}^2 = \sigma_1^2$$

$$E(Y_2) = \mu_{Y_2} = \mu_2 = \frac{1}{2}\ln\left[\frac{(\mu_{X_2})^4}{(\mu_{X_2})^2 + \sigma_{X_2}^2}\right]$$

$$Var(Y_2) = \sigma_{Y_2}^2 = \sigma_2^2 = \ln\left[\frac{(\mu_{X_2})^2 + \sigma_{X_2}^2}{(\mu_{X_2})^2}\right]$$

Assume that the pair

$$(X_1, X_2) \sim NLogN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$$

is a bivariate normal-lognormal distribution with density function

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \dfrac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho_{1,2}^2}\, x_2} e^{-\frac{1}{2}w} & -\infty < x_1 < \infty,\ 0 < x_2 < \infty \\ 0 \quad \text{otherwise} \end{cases}$$

where

$$\rho_{1,2} = \rho_{Y_1,Y_2} = \rho_{X_1, \ln X_2}$$

$$w = \frac{1}{1-\rho_{1,2}^2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{1,2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{\ln x_2-\mu_2}{\sigma_2}\right)+\left(\frac{\ln x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

Theorem A-1. If $(X_1,X_2) \sim NLogN((\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2,\rho_{1,2}))$ then

$$\rho_{1,2} = \rho_{X_1,X_2}\frac{(e^{\sigma_2^2}-1)^{1/2}}{\sigma_2}$$

Proof:

By definition

$$\rho_{X_1,X_2} = \frac{Cov(X_1,X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sigma_{X_1X_2}}{\sigma_{X_1}\sigma_{X_2}}$$

where

$$\sigma_{X_1X_2} = \int_0^{\infty}\int_{-\infty}^{\infty}(x_1-\mu_1)(x_2-\mu_2)f_{X_1,X_2}(x_1,x_2)dx_1dx_2$$

$$\sigma_{X_1} = \sigma_1$$

Since X_2 is lognormal, we have from table B-1 (appendix B)

$$\sigma_{X_2} = (e^{2\mu_2+\sigma_2^2}(e^{\sigma_2^2}-1))^{1/2} = E(X_2)(e^{\sigma_2^2}-1)^{1/2}$$

thus

$$\rho_{X_1,X_2} = \frac{\sigma_{X_1X_2}}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sigma_{X_1X_2}}{\sigma_1 E(X_2)(e^{\sigma_2^2}-1)^{1/2}}$$

To compute $\sigma_{X_1X_2}$ let

$$t_1 = \frac{x_1-\mu_1}{\sigma_1} \text{ and } t_2 = \frac{\ln x_2-\mu_2}{\sigma_2}$$

thus

$$\sigma_{X_1X_2} = \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1 t_1)(e^{\mu_2+\sigma_2 t_2}-\mu_2)e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_1^2-2\rho_{1,2}t_1t_2+t_2^2)}dt_1dt_2$$

$$= \frac{1}{2\pi\sqrt{1-\rho_{1,2}^2}}\int_{-\infty}^{\infty}(\sigma_1 t_1)[I_1-\mu_2 I_2]dt_1$$

where

$$I_1 = \int_{-\infty}^{\infty} e^{\mu_2+\sigma_2 t_2} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_1^2-2\rho_{1,2}t_1t_2+t_2^2)} dt_2$$

$$I_2 = \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_1^2-2\rho_{1,2}t_1t_2+t_2^2)} dt_2$$

To determine I_1 note that the integrand can be written as

$$I_1 = e^{\mu_2}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_1^2-2[\rho_{1,2}t_1+(1-\rho_{1,2}^2)\sigma_2]t_2+t_2^2)} dt_2$$

Letting

$$A = A(t_1) = \rho_{1,2}t_1 + (1-\rho_{1,2}^2)\sigma_2$$

and noting that

$$t_2^2 - 2At_2 = (t_2 - A)^2 - A^2$$

we can write

$$I_1 = e^{\mu_2} e^{-\frac{1}{2(1-\rho_{1,2}^2)}t_1^2} e^{\frac{1}{2(1-\rho_{1,2}^2)}A^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_2-A)^2} dt_2$$

$$I_1 = e^{\mu_2} e^{-\frac{1}{2(1-\rho_{1,2}^2)}t_1^2} e^{\frac{1}{2(1-\rho_{1,2}^2)}A^2} \sqrt{2\pi}\sqrt{(1-\rho_{1,2}^2)}$$

To determine I_2, note that the integrand can be written as

$$I_2 = e^{-\frac{1}{2(1-\rho_{1,2}^2)}t_1^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_2^2 - 2\rho_{1,2}t_1t_2)} dt_2$$

Letting

$$B = B(t_1) = \rho_{1,2}t_1$$

and noting that

$$t_2^2 - 2Bt_2 = (t_2 - B)^2 - B^2$$

we have

$$I_2 = e^{-\frac{1}{2(1-\rho_{1,2}^2)}t_1^2} e^{\frac{1}{2(1-\rho_{1,2}^2)}B^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho_{1,2}^2)}(t_2-B)^2} dt_2$$

$$I_2 = e^{-\frac{1}{2(1-\rho_{1,2}^2)}t_1^2} e^{\frac{1}{2(1-\rho_{1,2}^2)}B^2} \sqrt{2\pi}\sqrt{(1-\rho_{1,2}^2)}$$

Thus

$$I_1 - \mu_2 I_2 = e^{\frac{-t_1^2}{2(1-\rho_{1,2}^2)}} \sqrt{2\pi}\sqrt{(1-\rho_{1,2}^2)} \left[e^{\mu_2} e^{\frac{A^2}{2(1-\rho_{1,2}^2)}} - \mu_2 e^{\frac{B^2}{2(1-\rho_{1,2}^2)}} \right]$$

and

$$\sigma_{X_1X_2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma_1 t_1) e^{\frac{-t_1^2}{2(1-\rho_{1,2}^2)}} \left[e^{\mu_2} e^{\frac{A^2}{2(1-\rho_{1,2}^2)}} - \mu_2 e^{\frac{B^2}{2(1-\rho_{1,2}^2)}} \right] dt_1$$

$$\sigma_{X_1X_2} = \frac{1}{\sqrt{2\pi}} \left[e^{\mu_2}\sigma_1 \int_{-\infty}^{\infty} t_1 e^{\frac{-(t_1^2 - A^2)}{2(1-\rho_{1,2}^2)}} dt_1 - \mu_2 \sigma_1 \int_{-\infty}^{\infty} t_1 e^{\frac{-(t_1^2 - B^2)}{2(1-\rho_{1,2}^2)}} dt_1 \right]$$

$$\sigma_{X_1X_2} = \frac{1}{\sqrt{2\pi}} \left[e^{\mu_2}\sigma_1 \int_{-\infty}^{\infty} t_1 e^{-\frac{1}{2}(t_1 - \rho_{1,2}\sigma_2)^2 + \frac{1}{2}\sigma_2^2} dt_1 - \mu_2 \sigma_1 \int_{-\infty}^{\infty} t_1 e^{-t_1^2/2} dt_1 \right]$$

$$\sigma_{X_1X_2} = \frac{1}{\sqrt{2\pi}}\left[e^{\mu_2}\sigma_1 e^{\frac{1}{2}\sigma_2^2}\int_{-\infty}^{\infty} t_1 e^{-\frac{1}{2}(t_1-\rho_{1,2}\sigma_2)^2}dt_1 - \mu_2\sigma_1\cdot 0\right]$$

$$\sigma_{X_1X_2} = \frac{1}{\sqrt{2\pi}}\left[e^{\mu_2+\sigma_2^2/2}\sigma_1\rho_{1,2}\sigma_2\sqrt{2\pi}\right] = E(X_2)\rho_{1,2}\sigma_1\sigma_2$$

and

$$\rho_{X_1,X_2} = \frac{\sigma_{X_1X_2}}{\sigma_{X_1}\sigma_{X_2}} = \frac{E(X_2)\rho_{1,2}\sigma_1\sigma_2}{\sigma_1\left[e^{2\mu_2+\sigma_2^2}(e^{\sigma_2^2}-1)\right]^{1/2}} = \frac{E(X_2)\rho_{1,2}\sigma_1\sigma_2}{\sigma_1\left[E(X_2)(e^{\sigma_2^2}-1)^{1/2}\right]}$$

Thus

$$\rho_{1,2} = \rho_{X_1,X_2}\frac{(e^{\sigma_2^2}-1)^{1/2}}{\sigma_2} \qquad \text{(A-1)}$$

Theorem A-2. If $(X_1,X_2) \sim NLogN((\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2,\rho_{1,2}))$ then

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{1}{2}\left[(x_1-\mu_1)^2/\sigma_1^2\right]}$$

and

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2 x_2}e^{-\frac{1}{2}\left[(\ln x_2-\mu_2)^2/\sigma_2^2\right]}$$

Proof:

By definition

$$f_1(x_1) = \int_0^{\infty} f_{X_1,X_2}(x_1,x_2)dx_2$$

$$f_2(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)dx_1$$

The density function $f_{X_1,X_2}(x_1,x_2)$ can be factored as

$$f_{X_1,X_2}(x_1,x_2)=\left\{\frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x_1-\mu_1)^2/2\sigma_1^2}\right\}Q(x_1,x_2) \qquad \text{(A-2)}$$

where

$$Q(x_1,x_2)=\left\{\frac{1}{\sqrt{2\pi}(\sigma_2\sqrt{1-\rho_{1,2}^2})\,x_2}e^{-(\ln x_2-b)^2/2\sigma_2^2(1-\rho_{1,2}^2)}\right\}$$

and

$$b=\mu_2+\frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1)$$

Therefore

$$\begin{aligned} f_1(x_1) &= \int_0^\infty\left\{\frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x_1-\mu_1)^2/2\sigma_1^2}\right\}Q(x_1,x_2)dx_2 \\ &= \left\{\frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x_1-\mu_1)^2/2\sigma_1^2}\right\}\int_0^\infty Q(x_1,x_2)dx_2 \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{1}{2}\left[(x_1-\mu_1)^2/\sigma_1^2\right]} \end{aligned}$$

since the integrand is the probability density function of a $\mathrm{LogN}(b,\sigma_2^2(1-\rho_{1,2}^2))$ random variable.

To compute $f_2(x_2)$, the density function $f_{X_1,X_2}(x_1,x_2)$ is factored as

$$f_{X_1,X_2}(x_1,x_2)=Q^*(x_1,x_2)\left\{\frac{1}{\sqrt{2\pi}\sigma_2}\frac{1}{x_2}e^{-(\ln x_2-\mu_2)^2/2\sigma_2^2}\right\} \qquad \text{(A-3)}$$

where

$$Q^*(x_1,x_2)=\left\{\frac{1}{\sqrt{2\pi}(\sigma_1\sqrt{1-\rho_{1,2}^2})}e^{-(x_1-b^*)^2/2\sigma_1^2(1-\rho_{1,2}^2)}\right\}$$

and

$$b^*=\mu_1+\frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2-\mu_2)$$

Therefore

$$
\begin{aligned}
f_2(x_2) &= \int_{-\infty}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2} \right\} Q^*(x_1, x_2) dx_1 \\
&= \left\{ \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2} \right\} \int_{-\infty}^{\infty} Q^*(x_1, x_2) dx_1 \\
&= \frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2 / \sigma_2^2\right]}
\end{aligned}
$$

since the integrand is the probability density function of a $N(b^*, \sigma_1^2(1 - \rho_{1,2}^2))$ random variable.

Theorem A-3. If $(X_1, X_2) \sim NLogN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$X_1 | x_2 \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2} \rho_{1,2} (\ln x_2 - \mu_2), \sigma_1^2 (1 - \rho_{1,2}^2))$$

$$X_2 | x_1 \sim LogN(\mu_2 + \frac{\sigma_2}{\sigma_1} \rho_{1,2} (x_1 - \mu_1), \sigma_2^2 (1 - \rho_{1,2}^2))$$

Proof:

By definition, we have

$$f_{X_1|x_2}(x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_2(x_2)} = \frac{\left\{ \frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2 / \sigma_2^2\right]} \right\} Q^*(x_1, x_2)}{\frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2 / \sigma_2^2\right]}}$$

$$f_{X_1|x_2}(x_1) = Q^*(x_1, x_2)$$

thus, from A-3

$$X_1 | x_2 \sim N(b^*, \sigma_1^2 (1 - \rho_{1,2}^2))$$

where

$$b^* = \mu_1 + \frac{\sigma_1}{\sigma_2} \rho_{1,2} (\ln x_2 - \mu_2)$$

Similarly

$$f_{X_2|x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_1(x_1)} = \frac{\left\{\frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left[(x_1-\mu_1)^2/\sigma_1^2\right]}\right\} Q(x_1,x_2)}{\frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left[(x_1-\mu_1)^2/\sigma_1^2\right]}}$$

$$f_{X_2|x_1}(x_2) = Q(x_1,x_2)$$

thus, from A-2

$$X_2|x_1 \sim \mathrm{LogN}(b, \sigma_2^2(1-\rho_{1,2}^2))$$

where

$$b = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1 - \mu_1)$$

Theorem A-4. If $(X_1, X_2) \sim \mathrm{NLogN}((\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2,\rho_{1,2}))$ then

$$E(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma}\rho_{1,2}(x_1-\mu_1) + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)}$$

$$\mathrm{Var}(X_2|x_1) = e^{2(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1))} e^z(e^z - 1)$$

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2)$$

$$\mathrm{Var}(X_1|x_2) = \sigma_1^2(1-\rho_{1,2}^2)$$

where $z = \sigma_2^2(1-\rho_{1,2}^2)$

Proof:

Theorem A-3 proved that

$$X_2|x_1 \sim \mathrm{LogN}(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1), \sigma_2^2(1-\rho_{1,2}^2))$$

therefore, from table B-1 (appendix B)

$$E(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1) + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)}$$

$$Var(X_2|x_1) = e^{2(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1))} e^z(e^z - 1)$$

where $z = \sigma_2^2(1-\rho_{1,2}^2)$.

Theorem A-3 proved that

$$X_1|x_2 \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2))$$

therefore, it follows immediately from the properties of the normal that

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2)$$

and

$$Var(X_1|x_2) = \sigma_1^2(1-\rho_{1,2}^2)$$

Theorem A-5. If $(X_1, X_2) \sim NLogN((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$Median(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1)}$$

$$Mode(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1) - \sigma_2^2(1-\rho_{1,2}^2)}$$

$$Median(X_1|x_2) = E(X_1|x_2)$$

$$Mode(X_1|x_2) = E(X_1|x_2)$$

Proof:

Since $X_2|x_1$ is lognormally distributed (theorem A-3), from table B-1 (appendix B)

$$Median(X_2|x_1) = e^h = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1)}$$

and

$$\text{Mode}(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(x_1-\mu_1) - \sigma_2^2(1-\rho_{1,2}^2)}$$

Since $X_1|x_2$ is normally distributed (theorem A-3), it follows immediately that

$$\text{Median}(X_1|x_2) = E(X_1|x_2)$$

$$\text{Mode}(X_1|x_2) = E(X_1|x_2)$$

Property A-1. If $(X_1, X_2) \sim \text{NLogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$E(X_1|\text{Median}(X_2|\mu_1)) = \mu_1$$

Proof:

From theorem A-4 it was established that

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2)$$

From theorem A-5, we may write

$$\text{Median}(X_2|x_1 = \mu_1) = e^{\mu_2}$$

It follows that

$$\begin{aligned} E(X_1|\text{Median}(X_2|\mu_1)) = E(X_1|e^{\mu_2}) &= \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln e^{\mu_2} - \mu_2) \\ &= \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\mu_2 - \mu_2) \\ &= \mu_1 \end{aligned}$$

Appendix B

Theoretical Aspects of the LogNormal and Bivariate LogNormal Probability Distributions

The Univariate LogNormal

If X is a nonnegative random variable and $Y = \ln X$ has a normal distribution with

$$\mu_Y = E(Y) \quad \text{and} \quad \sigma_Y^2 = Var(Y)$$

then X is said to be lognormally distributed, that is

$$X \sim LogN(\mu_Y, \sigma_Y^2)$$

with density function

$$f_X(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma_Y x} e^{-\frac{1}{2}\left[(\ln x - \mu_Y)^2/\sigma_Y^2\right]} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B-1}$$

Equation B-1 is referred to as the lognormal distribution. Important properties of $f_X(x)$ are summarized in table B-1.

Table B-1 Properties of the Lognormal Distribution

Property	Expression
$\mu_X = E(X)$	$e^{\mu_Y + \frac{1}{2}\sigma_Y^2}$
$\sigma_X^2 = Var(X)$	$e^{2\mu_Y + \sigma_Y^2}(e^{\sigma_Y^2} - 1)$
Mode(X)	$e^{\mu_Y - \sigma_Y^2}$
Median(X)	e^{μ_Y}

Theorem B-1. If X is lognormally distributed with mean μ_X and variance σ_X^2 then

$$\mu_Y = \frac{1}{2}\ln\left[\frac{(\mu_X)^4}{(\mu_X)^2 + \sigma_X^2}\right]$$

and

$$\sigma_Y^2 = \ln\left[\frac{(\mu_X)^2 + \sigma_X^2}{(\mu_X)^2}\right]$$

The Bivariate LogNormal

Let $Y_1 = \ln X_1$ and $Y_2 = \ln X_2$ where X_1 and X_2 are random variables defined on $0 < x_1 < \infty$ and $0 < x_2 < \infty$. Define

$$E(Y_i) = \mu_{Y_i} = \mu_i \text{ and } \mathrm{Var}(Y_i) = \sigma_{Y_i}^2 = \sigma_i^2$$

for i = 1, 2. Suppose that Y_1 and Y_2 each have a normal distribution and that the pair

$$(X_1, X_2) \sim \text{Bivariate LogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$$

is a bivariate lognormal distribution with density function

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \dfrac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho_{1,2}^2}\,x_1 x_2} e^{-\frac{1}{2}w} & 0 < x_1 < \infty, 0 < x_2 < \infty \\ 0 \quad \text{otherwise} \end{cases} \qquad \text{(B-2)}$$

where

$$\rho_{1,2} = \rho_{Y_1,Y_2} = \rho_{\ln X_1, \ln X_2}$$

$$w = \frac{1}{1-\rho_{1,2}^2}\left\{\left(\frac{\ln x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho_{1,2}\left(\frac{\ln x_1 - \mu_1}{\sigma_1}\right)\left(\frac{\ln x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\ln x_2 - \mu_2}{\sigma_2}\right)^2\right\}$$

and

$$\mu_{Y_i} = \mu_i = \frac{1}{2}\ln\left[\frac{(\mu_{X_i})^4}{(\mu_{X_i})^2 + \sigma_{X_i}^2}\right] \quad \text{and} \quad \sigma_{Y_i}^2 = \sigma_i^2 = \ln\left[\frac{(\mu_{X_i})^2 + \sigma_{X_i}^2}{(\mu_{X_i})^2}\right]$$

for i = 1, 2.

Theorem B-2. If $(X_1, X_2) \sim \text{Bivariate LogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$\rho_{X_1,X_2} = \frac{e^{\rho_{1,2}\sigma_1\sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1}}$$

Proof:

By definition

$$\rho_{X_1,X_2} = \frac{Cov(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{E(X_1X_2) - E(X_1)E(X_2)}{\sigma_{X_1}\sigma_{X_2}} \tag{B-3}$$

Since $Y_1 = \ln X_1$ and $Y_2 = \ln X_2$ we can write

$$E(X_1X_2) = E(e^{Y_1}e^{Y_2}) = E(e^{Y_1+Y_2})$$

Recall that $Y_i \sim N(\mu_i, \sigma_i^2)$ for i = 1, 2. Thus, $E(e^{Y_1+Y_2})$ is recognized as a special evaluation of the moment generating function of a bivariate normal, which is

$$M(t_1, t_2) = E(e^{t_1Y_1+t_2Y_2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1y_1+t_2y_2} f(y_1, y_2)dy_1dy_2$$

$$= e^{(\mu_1t_1+\mu_2t_2)+\frac{1}{2}(\sigma_1^2t_1^2+2\rho_{Y_1,Y_2}\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2)}$$

for some real t_1 and t_2. Therefore,

$$E(X_1X_2) = E(e^{Y_1}e^{Y_2}) = E(e^{Y_1+Y_2}) = e^{(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+\sigma_2^2+2\rho_{Y_1,Y_2}\sigma_1\sigma_2)}$$

To determine the remaining terms in equation (B-3), we note that for $r \geq 0$ the moments of X are

$$E(X_i^r) = e^{r\mu_i+\frac{1}{2}r^2\sigma_i^2} \tag{B-4}$$

Thus,

$$E(X_1) = e^{\mu_1+\frac{1}{2}\sigma_1^2}$$
$$E(X_2) = e^{\mu_2+\frac{1}{2}\sigma_2^2}$$

$$\begin{aligned}
Var(X_1) &= E(X_1^2) - (E(X_1))^2 = e^{2\mu_1+2\sigma_1^2} - (e^{\mu_1+\frac{1}{2}\sigma_1^2})^2 \\
&= e^{2\mu_1+2\sigma_1^2} - e^{2\mu_1+\sigma_1^2} \\
Var(X_2) &= E(X_2^2) - (E(X_2))^2 = e^{2\mu_2+2\sigma_2^2} - (e^{\mu_2+\frac{1}{2}\sigma_2^2})^2 \\
&= e^{2\mu_2+2\sigma_2^2} - e^{2\mu_2+\sigma_2^2}
\end{aligned}$$

Substituting into equation B-3, we have

$$\rho_{X_1,X_2} = \frac{E(X_1X_2) - E(X_1)E(X_2)}{\sigma_{X_1}\sigma_{X_2}}$$

$$\rho_{X_1,X_2} = \frac{e^{(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+2\rho_{Y_1,Y_2}\sigma_1\sigma_2+\sigma_2^2)} - (e^{\mu_1+\frac{1}{2}\sigma_1^2})(e^{\mu_2+\frac{1}{2}\sigma_2^2})}{\sqrt{e^{2\mu_1+2\sigma_1^2} - e^{2\mu_1+\sigma_1^2}}\sqrt{e^{2\mu_2+2\sigma_2^2} - e^{2\mu_2+\sigma_2^2}}}$$

which can be factored as:

$$\rho_{X_1,X_2} = \frac{e^{(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+\sigma_2^2)}(e^{\rho_{1,2}\sigma_1\sigma_2} - 1)}{e^{(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+\sigma_2^2)}\sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1}}$$

where $\rho_{1,2} = \rho_{Y_1,Y_2} = \rho_{\ln X_1,\ln X_2}$

Thus,

$$\rho_{X_1,X_2} = \frac{e^{\rho_{1,2}\sigma_1\sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1}} \qquad \text{(B-5)}$$

which was to be shown.

Theorem B-3. If $(X_1, X_2) \sim$ Bivariate LogN$((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1\,x_1} e^{-\frac{1}{2}\left[(\ln x_1 - \mu_1)^2/\sigma_1^2\right]}$$

and

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2\,x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2/\sigma_2^2\right]}$$

Proof:

By definition

$$f_1(x_1) = \int_0^\infty f_{X_1,X_2}(x_1, x_2)\,dx_2$$

$$f_2(x_2) = \int_0^\infty f_{X_1,X_2}(x_1, x_2)\,dx_1$$

The density function $f_{X_1,X_2}(x_1, x_2)$ can be factored as

$$f_{X_1,X_2}(x_1, x_2) = \left\{ \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{x_1} e^{-(\ln x_1 - \mu_1)^2/2\sigma_1^2} \right\} Q(x_1, x_2) \qquad \text{(B-6)}$$

where

$$Q(x_1, x_2) = \left\{ \frac{1}{\sqrt{2\pi}(\sigma_2\sqrt{1 - \rho_{1,2}^2})\,x_2} e^{-(\ln x_2 - b)^2/2\sigma_2^2(1-\rho_{1,2}^2)} \right\}$$

and

$$b = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1)$$

Therefore

$$f_1(x_1) = \int_0^{\infty} \left\{ \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{x_1} e^{-(\ln x_1 - \mu_1)^2 / 2\sigma_1^2} \right\} Q(x_1, x_2) dx_2$$

$$= \left\{ \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{x_1} e^{-(\ln x_1 - \mu_1)^2 / 2\sigma_1^2} \right\} \int_0^{\infty} Q(x_1, x_2) dx_2$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1 x_1} e^{-\frac{1}{2}\left[(\ln x_1 - \mu_1)^2 / \sigma_1^2\right]}$$

since the integrand is the probability density function of a $\mathrm{LogN}(b, \sigma_2^2(1 - \rho_{1,2}^2))$ random variable.

To compute $f_2(x_2)$, the density function $f_{X_1,X_2}(x_1, x_2)$ is factored as

$$f_{X_1,X_2}(x_1, x_2) = Q^*(x_1, x_2) \left\{ \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2} \right\} \qquad \text{(B-7)}$$

where

$$Q^*(x_1, x_2) = \left\{ \frac{1}{\sqrt{2\pi}(\sigma_1 \sqrt{1 - \rho_{1,2}^2})\, x_1} e^{-(\ln x_1 - b^*)^2 / 2\sigma_1^2 (1 - \rho_{1,2}^2)} \right\}$$

and

$$b^* = \mu_1 + \frac{\sigma_1}{\sigma_2} \rho_{1,2} (\ln x_2 - \mu_2)$$

Therefore

$$f_2(x_2) = \int_0^{\infty} \left\{ \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2} \right\} Q^*(x_1, x_2) dx_1$$

$$= \left\{ \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2} \right\} \int_0^{\infty} Q^*(x_1, x_2) dx_1$$

$$= \frac{1}{\sqrt{2\pi}\sigma_2} \frac{1}{x_2} e^{-(\ln x_2 - \mu_2)^2 / 2\sigma_2^2}$$

since the integrand is the probability density function of a $\mathrm{LogN}(b^*, \sigma_1^2(1 - \rho_{1,2}^2))$ random variable.

Theorem B-4. If $(X_1, X_2) \sim$ Bivariate LogN$((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$X_1 | x_2 \sim \text{LogN}(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2))$$

$$X_2 | x_1 \sim \text{LogN}(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1), \sigma_2^2(1-\rho_{1,2}^2))$$

Proof:

By definition

$$f_{X_1|x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_2(x_2)} = \frac{\left\{\frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2/\sigma_2^2\right]}\right\} Q^*(x_1, x_2)}{\frac{1}{\sqrt{2\pi}\,\sigma_2 x_2} e^{-\frac{1}{2}\left[(\ln x_2 - \mu_2)^2/\sigma_2^2\right]}}$$

$$f_{X_1|x_2}(x_1) = Q^*(x_1, x_2)$$

thus, from B-7

$$X_1 | x_2 \sim \text{LogN}(b^*, \sigma_1^2(1-\rho_{1,2}^2))$$

where

$$b^* = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2)$$

Similarly

$$f_{X_2|x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_1(x_1)} = \frac{\left\{\frac{1}{\sqrt{2\pi}\,\sigma_1 x_1} e^{-\frac{1}{2}\left[(\ln x_1 - \mu_1)^2/\sigma_1^2\right]}\right\} Q(x_1, x_2)}{\frac{1}{\sqrt{2\pi}\,\sigma_1 x_1} e^{-\frac{1}{2}\left[(\ln x_1 - \mu_1)^2/\sigma_1^2\right]}}$$

$$f_{X_2|x_1}(x_2) = Q(x_1, x_2)$$

thus, from B-6

$$X_2 | x_1 \sim \text{LogN}(b, \sigma_2^2(1-\rho_{1,2}^2))$$

where

$$b = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1)$$

Theorem B-5. If $(X_1, X_2) \sim \text{Bivariate LogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$E(X_2|x_1) = x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1 + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)}$$

$$Var(X_2|x_1) = x_1^{2\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{2(\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1)} e^{z}(e^{z} - 1)$$

$$E(X_1|x_2) = x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2 + \frac{1}{2}\sigma_1^2(1-\rho_{1,2}^2)}$$

$$Var(X_1|x_2) = x_2^{2\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{2(\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2)} e^{z^*}(e^{z^*} - 1)$$

where

$$z = \sigma_2^2(1-\rho_{1,2}^2) \quad \text{and} \quad z^* = \sigma_1^2(1-\rho_{1,2}^2)$$

Proof:

Theorem B-4 proved that

$$X_2|x_1 \sim \text{LogN}(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1), \sigma_2^2(1-\rho_{1,2}^2))$$

Therefore, from table B-1

$$E(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1) + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)}$$

$$= x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1 + \frac{1}{2}\sigma_2^2(1-\rho_{1,2}^2)}$$

and

$$Var(X_2|x_1) = e^{2(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1))} e^{\sigma_2^2(1-\rho_{1,2}^2)}(e^{\sigma_2^2(1-\rho_{1,2}^2)} - 1)$$

$$= x_1^{2\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{2(\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1)} e^{z}(e^{z} - 1)$$

Theorem B-4 also proved that

$$X_1|x_2 \sim \mathrm{LogN}(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2))$$

Therefore, from table B-1

$$E(X_1|x_2) = e^{\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2) + \frac{1}{2}\sigma_1^2(1-\rho_{1,2}^2)}$$

$$= x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2 + \frac{1}{2}\sigma_1^2(1-\rho_{1,2}^2)}$$

and

$$\mathrm{Var}(X_1|x_2) = e^{2(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2))} e^{\sigma_1^2(1-\rho_{1,2}^2)}(e^{\sigma_1^2(1-\rho_{1,2}^2)} - 1)$$

$$= x_2^{2\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{2(\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2)} e^{z^*}(e^{z^*} - 1)$$

Theorem B-6. If $(X_1, X_2) \sim \mathrm{Bivariate\ LogN}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho_{1,2}))$ then

$$\mathrm{Median}(X_2|x_1) = x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1}$$

$$\mathrm{Mode}(X_2|x_1) = x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1 - \sigma_2^2(1-\rho_{1,2}^2)}$$

$$\mathrm{Median}(X_1|x_2) = x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2}$$

$$\mathrm{Mode}(X_1|x_2) = x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2 - \sigma_1^2(1-\rho_{1,2}^2)}$$

Proof:

From theorem B-4 and table B-1, it follows that

$$\mathrm{Median}(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1)} = x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1}$$

$$\mathrm{Mode}(X_2|x_1) = e^{\mu_2 + \frac{\sigma_2}{\sigma_1}\rho_{1,2}(\ln x_1 - \mu_1) - \sigma_2^2(1-\rho_{1,2}^2)}$$

$$= x_1^{\frac{\sigma_2}{\sigma_1}\rho_{1,2}} e^{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho_{1,2}\mu_1 - \sigma_2^2(1-\rho_{1,2}^2)}$$

$$\text{Median}(X_1|x_2) = e^{\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2)} = x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2}$$

$$\text{Mode}(X_1|x_2) = e^{\mu_1 + \frac{\sigma_1}{\sigma_2}\rho_{1,2}(\ln x_2 - \mu_2) - \sigma_1^2(1-\rho_{1,2}^2)}$$
$$= x_2^{\frac{\sigma_1}{\sigma_2}\rho_{1,2}} e^{\mu_1 - \frac{\sigma_1}{\sigma_2}\rho_{1,2}\mu_2 - \sigma_1^2(1-\rho_{1,2}^2)}$$

Property B-1. If $(X_1,X_2) \sim \text{Bivariate LogN}((\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2,\rho_{1,2}))$ then the conditional coefficient of dispersion D is

$$D_{F_{X_1|x_2}} = \frac{[\text{Var}(X_1|x_2)]^{1/2}}{E(X_1|x_2)} = \sqrt{(e^{z^*}-1)}$$
$$D_{F_{X_2|x_1}} = \frac{[\text{Var}(X_2|x_1)]^{1/2}}{E(X_2|x_1)} = \sqrt{(e^{z}-1)}$$

where $F_{X_1|x_2}$ and $F_{X_2|x_1}$ are the cumulative distributions of $f_{X_1|x_2}$ and $f_{X_2|x_1}$. This corollary is stated without proof since it directly follows from theorem B-5.